

# A Game-Theoretic Framework for Controlled Islanding in the Presence of Adversaries

Luyao Niu<sup>1†</sup>, Dinuka Sahabandu<sup>2†</sup>, Andrew Clark<sup>1</sup>, and Radha Poovendran<sup>2</sup>

<sup>1</sup> Department of Electrical and Computer Engineering,  
Worcester Polytechnic Institute, Worcester MA 01609, USA

<sup>2</sup> Network Security Lab, Department of Electrical and Computer Engineering,  
University of Washington, Seattle, WA 98195-2500, USA  
{lniu, aclark}@wpi.edu, {sdinuka, rp3}@uw.edu \*

**Abstract.** Controlled islanding effectively mitigates cascading failures by partitioning the power system into a set of disjoint islands. In this paper, we study the controlled islanding problem of a power system under disturbances introduced by a malicious adversary. We formulate the interaction between the grid operator and adversary using a game-theoretic framework. The grid operator first computes a controlled islanding strategy, along with the power generation for the post-islanding system to guarantee stability. The adversary observes the strategies of the grid operator. The adversary then identifies critical substations of the power system to compromise and trips the transmission lines that are connected with compromised substations. For our game formulation, we propose a double oracle algorithm based approach that solves the best response for each player. We show that the best responses for the grid operator and adversary can be formulated as mixed integer linear programs. In addition, the best response of the adversary is equivalent to a submodular maximization problem under a cardinality constraint, which can be approximated up to a  $(1 - \frac{1}{e})$  optimality bound in polynomial time. We compare the proposed approach with a baseline where the grid operator computes an islanding strategy by minimizing the power flow disruption without considering the possible response from the adversary. We evaluate both approaches using IEEE 9-bus, 14-bus, 30-bus, 39-bus, 57-bus, and 118-bus power system case study data. Our proposed approach achieves better performance than the baseline in about 44% of test cases, and on average it incurs about 12.27 MW less power flow disruption.

## 1 Introduction

The electric power system is a complex large-scale network that delivers electricity to customers. Modern power systems leverage Internet of Things (IoT) technologies and have integrated information and communication components [2], leading to the smart grid paradigm. However, incorporating cyber components

---

\* This work was supported by AFOSR grant FA9550-20-1-0074 and NSF grant CNS-1941670. † Authors contributed equally to this work.

exposes the power system to malicious cyber attacks [17, 20]. For example, the service outage incurred by the Ukrainian electric company in 2015 was caused by malicious cyber attacks [16].

Cyber attacks impact power systems by biasing the decision of the power grid operator, masking physical outages, and/or causing malfunctions of system components [29]. These disturbances can potentially lead to *cascading failures*. In a cascading failure, the outage of one component, e.g., a transmission line, shifts the load to other connected components, making them overload and fail. Power systems are under increasing risks of cascading failures since they are operated close to their capacity limits so as to meet ever-increasing electricity demands. Cascading failures can cause catastrophic economic consequences; the 2003 North American blackout, for example, left more than 55 million people in dark and caused 10 billion dollars loss [21]. An intelligent adversary may therefore take advantage of cascading failures to cause severe damage to power systems using limited resources.

Controlled islanding has been demonstrated to be an effective countermeasure against cascading failures [33]. Controlled islanding determines a subset of transmission lines to be tripped to partition the power system into multiple subsystems, following a disturbance such as transmission line outage.

Various techniques [7, 9, 18, 24, 28, 33] have been proposed for designing controlled islanding strategies with different criteria such as power flow disruption and power imbalance. To the best of our knowledge, however, there has been little study focusing on controlled islanding for power systems in the presence of malicious adversaries. Different from exogenous causes such as natural disasters and increasing load demand [32], intelligent adversaries can infer the islanding strategy of the grid operator and deliberately trip transmission lines to make the islands ineffective or even unstable. In the 2015 Ukrainian blackout, the adversaries compromised the substations and leveraged the strategies of the grid operator against the power system [16].

In this paper, we propose a game-theoretic model of controlled islanding to mitigate cascading failures in the presence of a malicious adversary. The adversary can compromise a subset of substations of the power system, and trip the transmission lines that are connected to the compromised substations. The grid operator aims at preventing cascading failure triggered by the adversary by implementing a controlled islanding strategy and designing corresponding post-islanding strategies. We make the following contributions:

- We model the interaction between the grid operator and adversary as a Stackelberg game, which we formulate as a mixed integer nonlinear program.
- We propose a double oracle algorithm based approach to solve for the strategies of the grid operator. The proposed approach iteratively computes the best response of each player.
- We analyze the best response for each player, and formulate it as a mixed integer linear program. In addition, we show the equivalence between the adversary’s best response and a submodular function maximization prob-

- lem with a cardinality constraint. A greedy algorithm can then be used to approximately compute the adversary's best response in polynomial time.
- We evaluate our proposed approach using IEEE 9-bus, 14-bus, 30-bus, 39-bus, 57-bus, and 118-bus power system case study data. We show that on average the power system performs better in 44% of the test cases and incurs 12.27 MW less power flow disruption when using the proposed approach, compared with a baseline that ignores the presence of an adversary.

The remainder of this paper is organized as follows. Related literature is reviewed in Section 2. Preliminary background on power system model and Stackelberg games is presented in Section 3. Section 4 gives the models for the adversary and grid operator, and maps the problem to a Stackelberg game. We present the solution approach in Section 5. Numerical evaluation results are presented in Section 6. We conclude the paper in Section 7.

## 2 Related Work

Computing controlled islanding strategies for power systems under large disturbances has been extensively studied. Typical approaches include slow coherency based islanding [33], ordered binary decision diagram (OBDD) methods [28], two-step spectral clustering technique [9], weak submodularity based controlled islanding [18], and mixed integer program based approaches [7, 24]. These works study the controlled islanding by assuming the disturbance has been detected and fixed. When there exists an adversary who can intelligently adjust its strategy, the islanding strategies computed using the aforementioned contributions need to be adjusted also to incorporate the possible response from the adversary.

Malicious attacks targeting power system have been reported and studied. The malicious attacks can be roughly classified into two categories. The first category of attacks manipulates the grid topology via transmission line switching [6, 22] and compromising substations [35]. Another category of attacks targets at the cyber components such as false data injection attack [8, 31]. In this paper, we consider a malicious adversary that compromises substations using cyber attacks and trips transmission lines. Different from existing literature, the adversary model studied in this paper not only identifies the critical components in the power system, but also considers the possible corrective action taken by the grid operator. In addition, existing topological models may discard the power system dynamics to simplify the computations [11], while this paper takes the physical properties of power systems into consideration.

In this paper, we map the interaction between the grid operator and adversary to a Stackelberg game. Stackelberg games have been widely used to model real-world security applications such as airport protection [26]. To compute the Stackelberg equilibrium [5] of the game in this paper, we propose a double oracle algorithm based approach [19]. Double oracle algorithm has been widely used to solve games of large-scale [3, 13, 14], due to the advantage that it avoids the enumeration over all possible strategies for the players.

### 3 Model and Preliminaries

#### 3.1 Power System Model

A power system with  $B$  substations and  $L$  transmission lines can be described by a graph  $\mathcal{G} = (\mathcal{B}, \mathcal{L})$ , where  $\mathcal{B} = \{1, \dots, B\}$  is the set of substations and  $\mathcal{L} \subseteq \mathcal{B} \times \mathcal{B}$  is the set of transmission lines. A transmission line  $l = (i, j) \in \mathcal{L}$  if substations  $i$  and  $j$  are connected via  $l$ . We define the set of neighboring substations for each  $i \in \mathcal{B}$  as  $\mathcal{T}(i) = \{j : (i, j) \in \mathcal{L}\}$ . The power injected to substation  $i$  is denoted as  $g_i$ , and the power drawn from substation  $i$  is denoted as  $d_i$ .

We consider DC power flow in the power system. The power flow  $P_{i,j}$  through each transmission line  $(i, j)$  is calculated as

$$P_{i,j} = S_{i,j}(\theta_i - \theta_j), \quad \forall (i, j) \in \mathcal{L} \quad (1)$$

where  $S_{i,j}$  is the electrical susceptance of transmission line  $(i, j)$ , and  $\theta_i, \theta_j$  are the voltage angles at substations  $i$  and  $j$ , respectively. Each substation  $i \in \mathcal{B}$  respects the power flow conservation law given as

$$\sum_{j \in \mathcal{T}(i)} P_{j,i} + g_i - d_i = 0, \quad \forall i \in \mathcal{B}. \quad (2)$$

Power generators exhibit varying behaviors following a large disturbance. Two generators are said to be *coherent* if their rotor angle deviations are within a certain tolerance [10]. To maintain the internal stability of the power system, the coherent generators need to be connected, while the non-coherent ones must be separated during islanding. In this paper, we assume that the set of power generators are classified into  $K$  coherent groups. Detailed techniques on computing coherent groups can be found in [4].

There are various metrics that have been proposed to measure the performance of power system when incurring disturbance. Typical metrics include power flow disruption [25, 30] and power imbalance [28, 34]. Minimum power flow disruption improves the transient stability of the system and reduces the risk of overloading transmission lines [12]. In this paper, we adopt power flow disruption as the performance metric, which is defined as

$$R(\mathcal{S}) = \sum_{(i,j) \in \mathcal{L}} (1 - z_{i,j}) \frac{|P_{i,j}| + |P_{j,i}|}{2}, \quad (3)$$

where  $\mathcal{S} \subseteq \mathcal{L}$  represents the set of tripped transmission lines. Parameter  $z_{i,j} = 1$  if  $(i, j) \in \mathcal{L} \setminus \mathcal{S}$  and  $z_{i,j} = 0$  if  $(i, j) \in \mathcal{S}$ .

#### 3.2 Stackelberg Game

Game theory models the interaction among multiple players. Consider a game consisting of two players, denoted as Player 1 and Player 2. Players 1 and 2 have their action spaces  $\mathcal{A}_1$  and  $\mathcal{A}_2$ , respectively. Each action of  $\mathcal{A}_1$  and  $\mathcal{A}_2$

is also known as the pure strategy for Player 1 and Player 2, respectively. A mixed strategy is a probability distribution over the action space. When Players 1 and 2 take strategies  $s_1$  and  $s_2$ , respectively, they obtain utilities  $U_1(s_1, s_2)$  and  $U_2(s_1, s_2)$  during the interaction.

Two-player Stackelberg games model interactions with information asymmetry, where Player 1 moves first by committing to a strategy, and Player 2 observes the strategy committed by Player 1 and chooses its strategy to maximize  $U_2(\cdot, \cdot)$ . Player 1 and Player 2 are also known as the leader and follower, respectively.

The solution concept of Stackelberg game is called Stackelberg equilibrium. We say strategies  $s_1^*$  and  $s_2^*$  for Players 1 and 2 comprise a Stackelberg equilibrium if  $s_1^* = \operatorname{argmax}_{s_1} U(s_1, s_2^*)$ , where  $s_2^* \in \mathcal{BR}(s_1^*)$  and  $\mathcal{BR}(s_1^*) = \operatorname{argmax}_{s_2} \{U_2(s_1^*, s_2)\}$  is the best response taken by Player 2 to  $s_1^*$ .

## 4 Problem Formulation

### 4.1 Adversary Model

We consider a power system  $\mathcal{G} = (\mathcal{B}, \mathcal{L})$ . A malicious adversary aims at destabilizing the power system and maximizing the power flow disruption. To achieve this goal, the adversary has two capabilities: (i) the adversary can compromise at most  $C$  substations  $\hat{\mathcal{B}} \subset \mathcal{B}$ , and (ii) the adversary can trip the set of transmission lines that are connected with the compromised substations. These capabilities have been demonstrated by real-world adversaries. For instance, the adversary that initiated the attack against the Ukrainian electric system compromised the substations and thus gained control over field devices [16].

We assume the adversary has access to the following information. The adversary knows the grid topology  $\mathcal{G} = (\mathcal{B}, \mathcal{L})$  and the power flow before it trips any transmission line. In addition, the adversary can observe the strategies taken by the grid operator (we will detail the model of grid operator in Section 4.2). We denote the information available to the adversary as  $\mathcal{I}^a$ . It has been reported that the adversary can harvest such information via cyber attacks [16].

In the following, we define the strategy for the adversary. A pure strategy for the adversary  $\tau : \mathcal{I}^a \rightarrow 2^{\mathcal{B}} \times 2^{\mathcal{L}}$  maps from the information set of the adversary to a pair of compromised substations  $\hat{\mathcal{B}}$  and tripped transmission lines  $\hat{\mathcal{L}}$ . A mixed strategy for the adversary  $\tau : \mathcal{I}^a \rightarrow \Delta(2^{\mathcal{B}} \times 2^{\mathcal{L}})$  maps from the information set of the adversary to a pair of probability distributions over  $2^{\mathcal{B}} \times 2^{\mathcal{L}}$ , where  $\Delta(\cdot)$  represents a probability distribution over some set. We define the set of proper adversary strategies as follows.

**Definition 1.** *We say strategy  $\tau$  is proper if the following conditions hold: (i)  $|\hat{\mathcal{B}}| \leq C$ , and (ii)  $\hat{\mathcal{L}} \subseteq \{(i, j) \in \mathcal{L} : i \in \hat{\mathcal{B}} \text{ or } j \in \hat{\mathcal{B}}\}$ .*

The adversary computes its strategy  $\tau$  as

$$\max_{\tau} R(\mathcal{S}) \quad (4a)$$

$$\text{subject to } \tau \text{ is proper} \quad (4b)$$

where  $R(S)$  is defined in Eqn. (3),  $S = \hat{\mathcal{L}} \cup \tilde{\mathcal{L}} \subseteq \mathcal{L}$ , and  $\tilde{\mathcal{L}}$  is the set of transmission lines tripped by the grid operator (we will introduce  $\tilde{\mathcal{L}}$  later in Section 4.2).

## 4.2 Grid Operator Model

In this subsection, we present the model of the grid operator. The goal of the grid operator is to protect the power system  $\mathcal{G} = (\mathcal{B}, \mathcal{L})$  when large disturbance is incurred. The grid operator has the following control capabilities. The grid operator can trip a subset of transmission lines  $\tilde{\mathcal{L}} \subset \mathcal{L} \setminus \hat{\mathcal{L}}$  to partition the power system into a collection of subsystems  $\{\mathcal{G}_k\}_{k=1}^K$ , where  $\mathcal{G}_k = (\mathcal{B}_k, \mathcal{L}_k)$ ,  $\mathcal{B}_k \subset \mathcal{B}$ , and  $\mathcal{L}_k \subset \mathcal{L}$ . A subsystem  $\mathcal{G}_k$  is also known as an island. After the power system is partitioned into subsystems, the grid operator controls the power injection  $g_i$  from each generator at each generation substation  $i \in \mathcal{B}$ .

We assume that the grid operator knows the grid topology and has perfect observation over the power system so that it can monitor the parameters such as the voltage angle at each substation, the power flow at each transmission line, the power injection from the generators, and the power drawn by the load demands. Additionally, the grid operator can compute the set of generator coherent groups. We denote the information available to the grid operator as  $\mathcal{I}^o$ .

A pure strategy for the grid operator is defined as  $\mu : \mathcal{I}^o \rightarrow 2^{\mathcal{L}} \times \mathbb{R}^N$  that maps from  $\mathcal{I}^o$  to the set of possibly tripped transmission lines and the space of power generations. Note that here the set of open transmission lines are tripped by the grid operator, and is different from those tripped by the adversary. A mixed strategy for the grid operator is defined as  $\mu : \mathcal{I}^o \rightarrow \Delta(2^{\mathcal{L}} \times \mathbb{R}^N)$ . We define the set of proper strategies for the grid operator.

**Definition 2.** *A strategy  $\mu$  for the grid operator is proper if the following conditions hold: (i) strategy  $\mu$  partitions the power system into disjoint subsystems, i.e.,  $\mathcal{B}_k \cap \mathcal{B}_{k'} = \emptyset$  and  $\mathcal{L}_k \cap \mathcal{L}_{k'} = \emptyset$ , (ii) the generators belonging to the same coherent group are within the same subsystem, (iii) each subsystem  $\mathcal{G}_k$  is connected, (iv) the post-islanding power generation and voltage angle are within the generation capacity for each generator and voltage angle bound for each substation, respectively, (v) the post-islanding power flow does not exceed the transmission capacities for all transmission lines, and (vi) the post-islanding power generation meets the load demand.*

## 4.3 Interaction Model Between the Grid Operator and Adversary

In this subsection, we present the interaction model between the grid operator and adversary. We denote the mixed strategy of the grid operator as  $\mu : \mathcal{I}^o \rightarrow \Delta(2^{\mathcal{L}}, \mathbb{R}^N)$ . The adversary observes strategy  $\mu$  of the grid operator by intruding into the power network and learning the strategies of the grid operator, and then computes a proper strategy  $\tau$ . Then the adversary executes its attack strategy  $\tau$  so as to destabilize the power system and maximize the power flow disruption. Once the grid operator detects the disturbance caused by the adversary, it

samples a pair  $(\tilde{\mathcal{L}}, g)$  following strategy  $\mu$ , and implements the sampled action to partition the system into a collection of subsystems.

There exists information asymmetry during the interaction between the grid operator and adversary. The adversary observes the strategy of the grid operator, while the grid operator has no information on the strategy of the adversary. This information asymmetry is captured by the Stackelberg game as described in Section 3.2. During this interaction, since the grid operator computes strategy  $\mu$  first, it becomes the leader in the Stackelberg game. The adversary, who observes the leader's strategy, is the follower in this setting.

The problem investigated in this paper is stated as follows.

*Problem 1.* Consider a power system  $\mathcal{G} = (\mathcal{B}, \mathcal{L})$ . Synthesize a proper strategy  $\mu$  for the grid operator that minimizes the power flow disruption, given that the adversary observes  $\mu$  and computes its best response to  $\mu$ , i.e.,

$$\min_{\mu} \mathbb{E}_{\mu}[R(\mathcal{S}(\mu, \tau))] \quad (5a)$$

$$\text{subject to } \mu \text{ is proper} \quad (5b)$$

$$\tau \in \mathcal{BR}(\mu) \quad (5c)$$

where  $\mathbb{E}_{\mu}[\cdot]$  denotes the expectation with respect to  $\mu$  and  $\mathcal{S}(\mu, \tau) = \hat{\mathcal{L}} \cup \tilde{\mathcal{L}} \subseteq \mathcal{L}$  is the set of tripped transmission lines that is jointly determined by  $\mu$  and  $\tau$ .

Note that the interaction between the grid operator and adversary is zero-sum. We can thus establish the existence of Stackelberg equilibrium strategies  $\mu$  and  $\tau$  of the game in Eqn. (5) using [5, Section 2].

## 5 Solution Approach

In this section, we present the solution approach to Problem 1. We prove that the sets of proper strategies  $\mu$  and  $\tau$  can be mapped to sets of mixed integer constraints. Then the optimization problem in Eqn. (5) is formulated as a mixed integer nonlinear program. We propose a double oracle algorithm based approach to compute the Stackelberg equilibrium strategies. The proposed approach computes the best response of each player in each iteration. We show that the best responses for both players can be formulated as mixed integer linear programs.

### 5.1 Mixed Integer Nonlinear Bi-level Optimization Formulation

In this subsection, we first map the set of proper strategies  $\mu$  and  $\tau$  to a set of mixed integer constraints. We then rewrite Eqn. (5) as a mixed integer nonlinear bi-level optimization problem.

We let  $y_i$  be a binary variable representing if substation  $i \in \mathcal{B}$  is compromised by the adversary ( $y_i = 1$ ) or not ( $y_i = 0$ ). We then have the following constraints:

$$\sum_{i \in \mathcal{B}} y_i \leq C, \quad y_i \in \{0, 1\}, \quad \forall i \in \mathcal{B}. \quad (6)$$

We define  $z_{i,j}^a$  as an indicator function for each transmission line  $(i, j)$  to represent if transmission line  $(i, j)$  is tripped ( $z_{i,j}^a = 0$ ) or not ( $z_{i,j}^a = 1$ ) by the adversary. Note that the adversary can trip a transmission line  $(i, j)$  if and only if substation  $i$  or  $j$  is compromised. We formulate this property as

$$z_{i,j}^a \in \{0, 1\}, \quad z_{i,j}^a = z_{j,i}^a, \quad \forall (i, j) \in \mathcal{L} \quad (7a)$$

$$z_{i,j}^a + y_i + y_j \geq 1, \quad \forall i \in \mathcal{B}, \quad \forall (i, j) \in \mathcal{L}. \quad (7b)$$

We denote  $y$  as the vector obtained by stacking  $y_i$  for all  $i \in \mathcal{B}$  and  $z^a$  as the vector obtained by stacking  $z_{i,j}^a$  for all  $(i, j) \in \mathcal{L}$ . We characterize the relations given by Eqn. (6) and (7) as follows.

**Lemma 1.** *The set of proper strategies for the adversary is equal to the set of feasible solutions  $(y, z^a)$  to Eqn. (6) and (7).*

*Proof.* We prove the statement using Definition 1. Consider condition (i) in Definition 1. Since  $y_i \in \{0, 1\}$  for all  $i \in \mathcal{B}$ , we have that if  $y$  is feasible to Eqn. (6), then at most  $C$  substations can be compromised. Consider condition (ii) in Definition 1. By Definition 1 and the definitions of  $z^a$  and  $y$ , we have that  $z_{i,j}^a = 0$  only if  $y_i + y_j \geq 1$ . However,  $y_i + y_j \geq 1$  does not necessarily imply that  $z_{i,j}^a = 0$ , i.e., the adversary can choose to not trip transmission line  $(i, j)$  even if substation  $i$  or  $j$  is compromised. In addition,  $z_{i,j}^a = 1$  must hold if  $y_i + y_j = 0$ . Summarizing these three possible scenarios, we have that  $z_{i,j}^a$ ,  $y_i$ , and  $y_j$  cannot be zero simultaneously, which is equivalent to Eqn. (7b). Combining the arguments above yields the lemma.  $\square$

Consider the set of proper strategies for the grid operator. Note that the pure strategy space for the grid operator grows exponentially with respect to the number of transmission lines  $L$ . To this end, we define a set of variables for each transmission line as a compact representation of the set of proper strategies.

Let  $x_{i,k}$  be a binary indicator representing if substation  $i$  is included in subsystem  $k$  ( $x_{i,k} = 1$ ) or not ( $x_{i,k} = 0$ ) for all  $i \in \mathcal{B}$  and  $k = 1, \dots, K$ . In addition, we let  $w_{i,j,k}$  be an indicator, representing if transmission line  $(i, j) \in \mathcal{L}$  is included ( $w_{i,j,k} = 1$ ) in subsystem  $\mathcal{G}_k = (\mathcal{B}_k, \mathcal{L}_k)$  or not ( $w_{i,j,k} = 0$ ). For each transmission line  $(i, j)$ , we define  $z_{i,j}^o$  as an indicator representing if transmission line  $(i, j) \in \mathcal{L}$  is tripped ( $z_{i,j}^o = 0$ ) or not ( $z_{i,j}^o = 1$ ) by the grid operator. We then formulate the constraints as

$$w_{i,j,k} \in \{0, 1\}, \quad w_{i,j,k} \leq x_{i,k}, \quad w_{i,j,k} \leq x_{j,k}, \quad \forall (i, j) \in \mathcal{L}, \quad \forall k = 1, \dots, K \quad (8a)$$

$$z_{i,j}^o = \sum_{k=1}^K w_{i,j,k}, \quad z_{i,j}^o \in \{0, 1\}, \quad z_{i,j}^o = z_{j,i}^o, \quad \forall (i, j) \in \mathcal{L} \quad (8b)$$

$$\sum_{k=1}^K x_{i,k} \leq 1, \quad \forall i \in \mathcal{B} \quad (8c)$$

$$x_{i,k} \in \{0, 1\}, \quad \forall i \in \mathcal{B}, \quad \forall k = 1, \dots, K \quad (8d)$$

$$z_{i,j}^o \leq z_{i,j}^a, \quad \forall (i, j) \in \mathcal{L} \quad (8e)$$



Eqn. (8e) captures the fact that the grid operator takes islanding action after the adversary executes the malicious attack. Hence, the grid operator cannot open a transmission line that has been tripped by the adversary.

Given the generator coherent groups, we let indicator  $v_{i,k} = 1$  if generation substation  $i$  is set as the reference generator and belongs to subsystem  $\mathcal{G}_k$ . Then using the coherent group, we can let

$$x_{j,k} = v_{i,k}, \forall i, j \in \mathcal{C}_k, \quad (9)$$

where  $\mathcal{C}_k$  represents the  $k$ -th generator coherent group. In addition, each subsystem  $\mathcal{G}_k$  is required to be connected. In order to incorporate this constraint, we define an auxiliary flow  $f_{i,j,k}$  on each transmission line  $(i, j)$  of subsystem  $k$ . Then the auxiliary flow should respect the flow conservation law given as

$$0 \leq f_{i,j,k} \leq Z z_{i,j}^o, \forall (i, j) \in \mathcal{L} \quad (10a)$$

$$v_{i,k} \sum_{j \in \mathcal{B}} x_{j,k} - x_{i,k} + \sum_{j \in \mathcal{T}(i)} f_{j,i,k} = \sum_{j \in \mathcal{T}(i)} f_{i,j,k}, \forall i \in \mathcal{B}, k = 1, \dots, K \quad (10b)$$

where  $Z$  is a sufficiently large positive constant. The first term of Eqn. (10b) implies that  $\sum_{j \in \mathcal{B}} x_{j,k}$  amount of auxiliary flow originates from the reference generator of subsystem  $\mathcal{G}_k$ . The second term of Eqn. (10b) indicates that one unit of auxiliary flow is consumed at substation  $i$ . The remaining two terms of Eqn. (10b) capture the incoming and outgoing auxiliary flows at substation  $i$ .

Relations given in Eqn. (8) to Eqn. (10) characterize the topological properties of each subsystem  $\mathcal{G}_k$ . In the following, we characterize the physical properties including the power flow and voltage angle in the power system after controlled islanding is implemented.

Each generator is constrained by its generation capacity modeled as

$$\underline{g}_i \leq g_i \leq \bar{g}_i, \forall i \in \mathcal{B} \quad (11)$$

where  $\underline{g}_i$  and  $\bar{g}_i$  are the minimum and maximum power generation capacities for generation substation  $i$ , respectively. We denote the post-islanding power flow on transmission line as  $\tilde{P}_{i,j}$  and voltage angle of substation  $i$  as  $\theta_i$ . By Eqn. (1), we have that  $S_{i,j}(\theta_i - \theta_j) - \tilde{P}_{i,j} = 0$  holds for all  $(i, j) \in \mathcal{L}$ . To incorporate the fact that the transmission line  $(i, j)$  can be tripped by the grid operator and adversary, we have that

$$-(1 - z_{i,j}^o z_{i,j}^a)Z \leq S_{i,j}(\theta_i - \theta_j) - \tilde{P}_{i,j} \leq (1 - z_{i,j}^o z_{i,j}^a)Z, \quad (12)$$

where  $Z$  is a sufficiently large positive constant. Taking the transmission line capacity and voltage angle bound into consideration, we have

$$\underline{P}_{i,j} z_{i,j}^o z_{i,j}^a \leq \tilde{P}_{i,j} \leq \bar{P}_{i,j} z_{i,j}^o z_{i,j}^a, \forall (i, j) \in \mathcal{L}, \underline{\theta}_i \leq \theta_i \leq \bar{\theta}_i, \forall i \in \mathcal{B}, \quad (13)$$

where  $\bar{P}_{i,j}$  and  $\underline{P}_{i,j}$  are the maximal and minimal power flow capacity for transmission line  $(i, j)$ , and  $\underline{\theta}_i$  and  $\bar{\theta}_i$  are respectively the minimum and maximum

voltage angle at substation  $i$ . Using Eqn. (13), we observe that the only feasible power flow through a tripped transmission line is zero. By Eqn. (2), the power balance at each substation  $i$  is modeled as

$$\sum_{j \in \mathcal{T}(i)} \tilde{P}_{j,i} + g_i - d_i = 0, \forall i \in \mathcal{B}. \quad (14)$$

We denote  $w, x, v, z^o, f, \tilde{P}, g$ , and  $\theta$  as the vectors or matrices that are obtained by stacking  $w_{i,j,k}, x_{i,k}, v_{i,k}, z_{i,j}^o, f_{i,j,k}, \tilde{P}_{i,j}, g_i$ , and  $\theta_i$ , respectively. We characterize Eqn. (8) to (14) as follows.

**Lemma 2.** *If variables  $w, x, t, z^o, f, \tilde{P}, g$ , and  $\theta$  are feasible to Eqn. (8) to Eqn. (14), then these variables represent a proper strategy for the grid operator as given in Definition 2*

*Proof.* Consider variables  $w, x, t, z^o, f, \tilde{P}, g$ , and  $\theta$  that are feasible to Eqn. (8) to Eqn. (14). We then verify that conditions (i)-(vi) in Definition 2 are satisfied.

Satisfaction of Condition (i). Suppose that Eqn. (8) is satisfied while the subsystems  $\mathcal{G}_k$  are not disjoint. Thus we have that there exists  $k \neq k'$  such that  $x_{i,k} = x_{i,k'} = 1$  holds for some  $i \in \mathcal{B}$  or  $w_{i,j,k} = w_{i,j,k'} = 1$  holds for some  $(i,j) \in \mathcal{L}$ . If  $x_{i,k} = x_{i,k'} = 1$  holds for some  $i \in \mathcal{B}$ , then Eqn. (8c) is violated.  $w_{i,j,k} = w_{i,j,k'} = 1$  holds for some  $(i,j) \in \mathcal{L}$ , then Eqn. (8b) implies that  $z_{i,j}^o > 1$ , which leads to contradiction. Thus, condition (i) of Definition 2 is satisfied.

Satisfaction of condition (ii). Condition (ii) holds immediately by the definition of  $x_{i,k}, v_{i,k}$ , and Eqn. (9).

Satisfaction of condition (iii). Suppose  $f$  satisfies Eqn. (10) while there exists some subsystem  $\mathcal{G}_k$  that is not connected. Without loss of generality, we assume that substation  $i$  belonging to subsystem  $\mathcal{G}_k$  is not connected with substations  $j \in \mathcal{B}_k \setminus \{i\}$ . If substation  $i$  is not the  $k$ -th reference generation substation, then Eqn. (10b) becomes  $-x_{i,k} = 0$ , which contradicts our hypothesis that  $x_{i,k} = 1$ . If substation  $i$  is the  $k$ -th reference generation substation, then  $v_{i,k} = 1$  and Eqn. (10b) is rewritten as  $\sum_{j \in \mathcal{B}} x_{j,k} - x_{i,k} = 0$ , which leads contradiction since there exists  $j \in \mathcal{B}_k \setminus \{i\}$  such that  $x_{j,k} = 1$ . Therefore, we can conclude that condition (iii) of Definition 2 is satisfied when Eqn. (10) is satisfied.

Satisfaction of condition (iv). Condition (iv) follows from Eqn. (11) and (13).

Satisfaction of condition (v). Consider Eqn. (13) for a transmission line  $(i,j)$ . If transmission line  $(i,j)$  is tripped by either the adversary or the grid operator, then Eqn. (13) implies that  $\tilde{P}_{i,j} = 0$ , which satisfies the power flow equation. If transmission line  $(i,j)$  is tripped by neither the adversary nor the grid operator, then power flow  $\tilde{P}_{i,j}$  satisfies Eqn. (1). The transmission line capacity constraint then immediately follows from Eqn. (13).

Satisfaction of condition (vi). Condition (vi) of Definition 2 holds by the definitions of  $\tilde{P}_{i,j}, g_i, d_i$ , and Eqn. (2).  $\square$

Lemma 1 and 2 imply that we can represent the pure strategy space using a collection of variables, whose size is polynomial in terms of  $B$  and  $L$ . Using

these variables, we can rewrite optimization problem (5) as

$$\min_{w,x,z^o,f,\tilde{P},g,\theta} \mathbb{E}_\mu \left[ \sum_{(i,j) \in \mathcal{L}} (1 - z_{i,j}^o z_{i,j}^a) \frac{|P_{i,j}| + |P_{j,i}|}{2} \right] \quad (15a)$$

$$\text{subject to Eqn. (8) to Eqn. (14)} \quad (15b)$$

$$(y, z^a) \in \operatorname{argmax}_{(i,j) \in \mathcal{L}} \sum (1 - z_{i,j}^o z_{i,j}^a) \frac{|P_{i,j}| + |P_{j,i}|}{2} \quad (15c)$$

$$\text{subject to Eqn. (6) to Eqn. (7)} \quad (15d)$$

Eqn. (15a) to (15b) and Eqn. (15c) to (15d) are known as the upper and lower level of bi-level optimization program (15), respectively. We remark that although  $\tilde{P}$  and  $\theta$  are set as decision variables in optimization program (15), they are inherently determined once the grid topology and power generation  $g$  are given. Therefore, the upper level of Eqn. (15) is interpreted as computing the partitions of the power system using  $z^o$  as a corrective measure against the malicious attack. For the power system partition  $z^o$ , the grid operator needs to compute power generation  $g$  so that there exists some feasible post-islanding DC power flow  $\tilde{P}_{i,j}$  satisfies conditions (iv)-(vi) in Definition 2.

## 5.2 Double Oracle Algorithm Based Approach

In this subsection, we present a double oracle algorithm based approach to solve Problem 1. The proposed approach alternatively solves the upper and lower level of optimization problem (15), and converges to the Stackelberg equilibrium.

---

### Algorithm 1 Double Oracle Algorithm for Controlled Islanding

---

- 1: Initialize a set of actions  $(\mathcal{Z}^o, G)$  for the grid operator, with each  $(z^o, g) \in (\mathcal{Z}^o, G)$  being feasible to Eqn. (8) to Eqn. (14)
  - 2: Initialize a set of actions  $(\mathcal{Y}, \mathcal{Z}^a)$  for the adversary, with each  $(y, z^a) \in (\mathcal{Y}, \mathcal{Z}^a)$  being feasible to Eqn. (6) to Eqn. (7)
  - 3: **while** not converge **do**
  - 4:   Solve for  $(\mu, \tau)$  by constraining the grid operator and adversary to take actions from  $(\mathcal{Z}^o, G)$  and  $(\mathcal{Y}, \mathcal{Z}^a)$ , respectively
  - 5:   Compute  $(z^o, g)$ , assuming the adversary takes strategy  $\tau$
  - 6:    $(\mathcal{Z}^o, G) \leftarrow (\mathcal{Z}^o, G) \cup (z^o, g)$
  - 7:   Solve for  $(y, z^a)$ , assuming the grid operator takes strategy  $\mu$
  - 8:    $(\mathcal{Y}, \mathcal{Z}^a) \leftarrow (\mathcal{Y}, \mathcal{Z}^a) \cup (y, z^a)$
  - 9: **end while**
  - 10: **return**  $(\mu, \tau)$
-

Algorithm 1 presents the double oracle approach. It consists of four steps. The first step is presented in lines 1 to 2 of Algorithm 1. In this step, the algorithm initializes a set of pure strategies for the grid operator and adversary, respectively. The initialized pure strategies are proper. The second step corresponds to line 4 of Algorithm 1. In this step, the algorithm solves a mixed strategy  $\mu$  for the grid operator and a pure strategy  $\tau$  for the adversary. The reason that pure strategy is considered for the adversary is that it is the follower in the game, whose pure strategies suffice for best response calculation [5]. Note that here mixed strategy  $\mu$  defines a probability distribution over  $(\mathcal{Z}^o, G)$ , rather than the full strategy space. Similarly, best response  $\tau$  gives an action selected from  $(\mathcal{Y}, \mathcal{Z}^a)$ . Lines 5 to 6 correspond to the third step of Algorithm 1. This step computes a pure strategy for the grid operator over all the feasible strategies, given that the adversary plays strategy  $\tau$ . The fourth step is presented in lines 7-8 of Algorithm 1, where the adversary computes its best response to mixed strategy  $\mu$ . The second to the last step of Algorithm 1 are executed in an iterative manner. The iteration terminates when no pure strategies for the grid operator and adversary are included in line 6 and line 8. The worst-case number of iterations Algorithm 1 can take to converge is  $(2^L - 1)$ , which is identical to solving for the Stackelberg equilibrium using linear program [5]. However, implementing the linear program requires constructing the action spaces of dimensions  $2^L$  for the grid operator and adversary and the corresponding constraints.

Given the current set of pure strategies  $(\mathcal{Z}^o, G)$  and  $(\mathcal{Y}, \mathcal{Z}^a)$  for the grid operator and adversary, respectively, line 4 of Algorithm 1 can be formulated as

$$\min_{\mu, \tau, r} \sum_{z^o \in \mathcal{Z}^o} \sum_{z^a \in \mathcal{Z}^a} \mu(z^o) \tau(z^a) \sum_{(i,j) \in \mathcal{L}} \frac{1 - z_{i,j}^o z_{i,j}^a}{2} (|P_{i,j}| + |P_{j,i}|) \quad (16a)$$

$$\text{subject to} \quad \sum_{z^o \in \mathcal{Z}^o} \mu(z^o) = 1 \quad (16b)$$

$$\mu(z^o) \in [0, 1], \quad \forall z^o \in \mathcal{Z}^o \quad (16c)$$

$$\sum_{z^a \in \mathcal{Z}^a} \tau(z^a) = 1 \quad (16d)$$

$$\tau(z^a) \in \{0, 1\}, \quad \forall z^a \in \mathcal{Z}^a \quad (16e)$$

$$\begin{aligned} 0 \leq r - \sum_{z^o \in \mathcal{Z}^o} \mu(z^o) \sum_{(i,j) \in \mathcal{L}} \frac{1 - z_{i,j}^o z_{i,j}^a}{2} (|P_{i,j}| + |P_{j,i}|) \\ \leq (1 - \tau(z^a))Z, \quad \forall z^a \in \mathcal{Z}^a \end{aligned} \quad (16f)$$

$$r \geq 0 \quad (16g)$$

$$\text{Eqn. (6) to Eqn. (14)} \quad (16h)$$

where  $Z$  is a sufficiently large positive constant. Optimization problem (16) slightly abuses the notation, and uses  $\mu(z^o)$  and  $\tau(z^a)$  to represent the probabilities the grid operator applies  $z^o$  and the adversary applies  $z^a$ , respectively. Constraints (16b) and (16c) ensures that  $\mu$  is a well-defined mixed strategy. Con-

straints (16d) and (16e) capture the fact that the adversary computes a pure strategy as its best response. Constraints (16f) and (16g) quantify the optimal power flow disruption  $r$  that the adversary can cause. By Eqn. (16f), we have that if the adversary plays its best response ( $\tau(z^a) = 1$ ), then it can achieve  $r$  amount of power flow disruption. For  $\tau(z^a) = 0$ , constraint (16f) is satisfied trivially. Constraint (16h) guarantees that the strategies are proper.

Optimization problem (16) is a mixed integer nonlinear program (MINLP). The nonlinearity can be mitigated by defining a new variable  $u_{z^o z^a}$ , which is defined as  $u_{z^o z^a} = \mu(z^o)\tau(z^a)$  for all  $z^o, z^a$  satisfying Eqn. (16h) and  $u_{z^o z^a} = 0$  otherwise. The constraints defined on  $u_{z^o z^a}$  are

$$u_{z^o z^a} \in [0, 1], \forall z^o \in \mathcal{Z}^o, \forall z^a \in \mathcal{Z}^a, 0 \leq \sum_{z^a \in \mathcal{Z}^a} u_{z^o z^a} \leq 1, \forall z^o \in \mathcal{Z}^o \quad (17a)$$

$$\sum_{z^o \in \mathcal{Z}^o} \sum_{z^a \in \mathcal{Z}^a} u_{z^o z^a} = 1. \quad (17b)$$

Using  $u_{z^o z^a}$ , MINLP (16) is converted to a mixed integer linear program (MILP):

$$\min_{u, \tau, r} \sum_{z^o \in \mathcal{Z}^o} \sum_{z^a \in \mathcal{Z}^a} u_{z^o z^a} \sum_{(i,j) \in \mathcal{L}} \frac{1 - z_{i,j}^o z_{i,j}^a}{2} (|P_{i,j}| + |P_{j,i}|) \quad (18a)$$

$$\text{subject to} \quad \text{Eqn. (16d), (16e), (16g), and (17)} \quad (18b)$$

$$0 \leq r - \sum_{z^o \in \mathcal{Z}^o} \left[ \sum_{(i,j) \in \mathcal{L}} \frac{1 - z_{i,j}^o z_{i,j}^a}{2} (|P_{i,j}| + |P_{j,i}|) \right] \left[ \sum_{\bar{z}^a \in \mathcal{Z}^a} u_{z^o \bar{z}^a} \right] \leq (1 - \mathbb{P}(z^a))Z, \forall z^a \in \mathcal{Z}^a \quad (18c)$$

Similar techniques for converting MINLP to MILP have been used in [23]. The equivalence between MILP (18) and MINLP (16) is presented as follows.

**Lemma 3.** *The MINLP (16) is equivalent to the MILP (18).*

*Proof.* We first prove that the objective functions of MINLP (16) and MILP (18) are identical. We then show that a feasible solution to Eqn. (16) is also feasible to Eqn. (18), and vice versa. The equivalence between (16a) and (18a) holds by the construction of  $u_{z^o z^a}$ .

Let  $\mu$ ,  $\tau$ , and  $r$  be feasible solutions to Eqn. (16). Let  $u_{z^o z^a} = \mu(z^o)\tau(z^a)$ . We have that  $u_{z^o z^a} \in [0, 1]$  holds by the construction of  $u_{z^o z^a}$ . By the definition of  $u_{z^o z^a}$ , we have that  $\sum_{z^o \in \mathcal{Z}^o} \sum_{z^a \in \mathcal{Z}^a} u_{z^o z^a} = 1$  holds by constraints (16b) and (16e). Inequality  $0 \leq \sum_{z^a \in \mathcal{Z}^a} u_{z^o z^a} \leq 1$  holds by Eqn. (16d) and the definition of  $u_{z^o z^a}$ . Constraint (18c) follows by substituting  $u_{z^o z^a}$  into Eqn. (16f).

Let  $u$ ,  $\tau$ , and  $r$  be feasible to Eqn. (18). We prove that  $\mu$ ,  $\tau$ , and  $r$  are feasible solutions to Eqn. (16), where  $\mu(z^o) = \sum_{z^a} u_{z^o z^a}$ . Since  $\mu(z^o) = \sum_{z^a} u_{z^o z^a}$  and  $\tau(z^a) \in \{0, 1\}$ , we have that constraint  $\sum_{z^o \in \mathcal{Z}^o} \sum_{z^a \in \mathcal{Z}^a} u_{z^o z^a} = 1$  implies that constraint (16b) holds. Using  $\mu(z^o) = \sum_{z^a} u_{z^o z^a}$ , constraint  $0 \leq \sum_{z^a \in \mathcal{Z}^a} u_{z^o z^a} \leq 1$  can be rewritten as  $0 \leq \mu(z^o) \leq 1$ , i.e., constraint (16c). Similarly, the equivalence between constraints (18c) and (16f) follows by  $\mu(z^o) = \sum_{z^a \in \mathcal{Z}^a} u_{z^o z^a}$ .  $\square$

Consider line 5 of Algorithm 1. This corresponds to Eqn. (15a) to (15b) when the strategy of the adversary is fixed. Given any feasible  $(y, z^a)$  for the adversary, the grid operator solves the following optimization problem:

$$\min_{w, x, z^o, f, \bar{P}, g, \theta} \sum_{(i,j) \in \mathcal{L}} (1 - z_{i,j}^o z_{i,j}^a) \frac{|P_{i,j}| + |P_{j,i}|}{2} \quad (19a)$$

$$\text{subject to Eqn. (8) to Eqn. (14)} \quad (19b)$$

Eqn. (19) is an MILP and can be solved using commercial solvers. Note that optimization problem (19) computes a pure strategy  $(z^o, g)$  for the grid operator.

In the following, we present an MILP for line 7 of Algorithm 1, which corresponds to solving Eqn. (15c) to (15d) when the strategy of the grid operator is given. Since the grid operator plays a mixed strategy, the goal of the adversary then becomes maximizing the expected power flow disruption, where the expectation is taken over mixed strategy  $\mu$ . With a slight abuse of notation, we denote the probability that the grid operator trips the transmission lines corresponding to  $z^o$  as  $\mu(z^o)$ . The MILP corresponding to line 7 of Algorithm 1 is given as

$$\max_{y, z^a} \sum_{z^o \in \mathcal{Z}^o} \mu(z^o) \sum_{(i,j) \in \mathcal{L}} (1 - z_{i,j}^o z_{i,j}^a) \frac{|P_{i,j}| + |P_{j,i}|}{2} \quad (20a)$$

$$\text{subject to Eqn. (6) to Eqn. (7)} \quad (20b)$$

In the following, we show that the optimization problem (20) can be mapped to a submodular maximization problem subject to a cardinality constraint. As a consequence, a greedy algorithm is presented to solve for a pure strategy for the adversary. We relax optimization problem (20) as

$$\max_{\hat{\mathcal{B}}} \sum_{z^o \in \mathcal{Z}^o} \mu(z^o) \left[ \sum_{(i,j) \in \hat{\mathcal{L}}} \frac{|P_{i,j}| + |P_{j,i}|}{2} + \sum_{(i,j) \in \tilde{\mathcal{L}}} \frac{|P_{i,j}| + |P_{j,i}|}{2} \right] \quad (21a)$$

$$\text{subject to } \hat{\mathcal{L}} = \{(i, j) \in \mathcal{L} : i \in \hat{\mathcal{B}} \text{ or } j \in \hat{\mathcal{B}}\} \quad (21b)$$

$$|\hat{\mathcal{B}}| \leq C \quad (21c)$$

where  $\tilde{\mathcal{L}} \subset \mathcal{L}$  is the set of transmission lines tripped by the grid operator when taking action  $z^o$ . We characterize the relation between optimization problem (20) and (21) using the following lemma.

**Lemma 4.** *Given the strategy of the grid operator, the optimal solution to optimization problem (21) is identical to that of optimization problem (20).*

*Proof.* We omit the proof due to space constraint.  $\square$

We now map optimization problem (21) to a problem of maximizing a submodular function subject to a cardinality constraint. We define  $\chi_{i,j}(\hat{\mathcal{B}})$  as

$$\chi_{i,j}(\hat{\mathcal{B}}) = \begin{cases} 1 & \text{if } i \in \hat{\mathcal{B}} \text{ or } j \in \hat{\mathcal{B}} \\ 0 & \text{otherwise} \end{cases} \quad (22)$$

Using the definition of  $\chi_{i,j}(\hat{\mathcal{B}})$ , optimization problem (21) can be rewritten as

$$\max_{\hat{\mathcal{B}}} \sum_{z^o \in \mathcal{Z}^o} \mu(z^o) \sum_{(i,j) \in \mathcal{L}} \chi_{i,j}(\hat{\mathcal{B}}) \frac{|P_{i,j}| + |P_{j,i}|}{2} \quad (23a)$$

$$\text{subject to } |\hat{\mathcal{B}}| \leq C \quad (23b)$$

We have the following result.

**Proposition 1.** *Objective function (23a) is submodular and nondecreasing with respect to  $\hat{\mathcal{B}}$ .*

*Proof.* We first prove Eqn. (23a) is submodular with respect to  $\hat{\mathcal{B}}$  using the definition of submodularity. Let  $\hat{\mathcal{B}}_2 \subseteq \hat{\mathcal{B}}_1 \subseteq \mathcal{B}$ . By Eqn. (22), we have that

$$\chi_{i,j}(\hat{\mathcal{B}} \cup \{h\}) - \chi_{i,j}(\hat{\mathcal{B}}) = \begin{cases} 1 & \text{if } h = i \text{ or } h = j \text{ and } i, j \notin \hat{\mathcal{B}}, \forall \hat{\mathcal{B}} \subseteq \mathcal{B} \\ 0 & \text{otherwise} \end{cases} \quad (24)$$

Suppose that  $\chi_{i,j}(\hat{\mathcal{B}}_1 \cup \{h\}) - \chi_{i,j}(\hat{\mathcal{B}}_1) = 1$ . Since  $\hat{\mathcal{B}}_2 \subseteq \hat{\mathcal{B}}_1$ , we have that  $h = i$  or  $h = j$  and  $i, j \notin \hat{\mathcal{B}}_2$ . Then we have that  $\chi_{i,j}(\hat{\mathcal{B}}_2 \cup \{h\}) - \chi_{i,j}(\hat{\mathcal{B}}_2) = 1$ . Therefore, we have that  $\chi_{i,j}(\hat{\mathcal{B}}_2 \cup \{h\}) - \chi_{i,j}(\hat{\mathcal{B}}_2) \geq \chi_{i,j}(\hat{\mathcal{B}}_1 \cup \{h\}) - \chi_{i,j}(\hat{\mathcal{B}}_1)$  holds for all  $\hat{\mathcal{B}}_2 \subseteq \hat{\mathcal{B}}_1 \subseteq \hat{\mathcal{B}}$ , which implies that Eqn. (22) is submodular with respect to  $\hat{\mathcal{B}}$ .

Consider  $\hat{\mathcal{B}}_2 \subset \hat{\mathcal{B}}_1$ . Then there must exist some  $i \in \hat{\mathcal{B}}_1$  while  $i \notin \hat{\mathcal{B}}_2$ . Let  $j \in \mathcal{B}$  be a substation satisfying  $j \notin \hat{\mathcal{B}}_1$ . Using Eqn. (22), we have that  $\chi_{i,j}(\hat{\mathcal{B}}_2) = 0 \leq \chi_{i,j}(\hat{\mathcal{B}}_1) = 1$ . Let  $j \in \mathcal{B}$  be a substation satisfying  $j \in \hat{\mathcal{B}}_2$ . Then we have that  $\chi_{i,j}(\hat{\mathcal{B}}_2) = \chi_{i,j}(\hat{\mathcal{B}}_1) = 0$ . If  $j \in \hat{\mathcal{B}}_1$  holds while  $j \notin \hat{\mathcal{B}}_2$  does not hold. Then we have that  $\chi_{i,j}(\hat{\mathcal{B}}_2) = 0 \leq \chi_{i,j}(\hat{\mathcal{B}}_1) = 1$ . Summarizing the arguments above, we have that Eqn. (22) is nondecreasing with respect to  $\hat{\mathcal{B}}$ .

Combining the arguments above, we have that Eqn. (23a) is a summation of non-negative submodular and nondecreasing functions. Therefore, Eqn. (23a) is a submodular and nondecreasing function with respect to  $\hat{\mathcal{B}}$ .  $\square$

According to Proposition 1, optimization problem (20) is equivalent to a submodular maximization problem with cardinality constraint. Optimization problem (23) can be solved using a greedy algorithm in polynomial time [15]. It has been shown that the greedy algorithm achieves  $1 - \frac{1}{e}$  optimality guarantee [15].

We conclude this section by giving the convergence and optimality of double oracle algorithm [19]. We state the result in the following lemma.

**Lemma 5.** *Algorithm 1 converges to the Stackelberg equilibrium within finitely many iterations if the best responses in line 5 and line 7 are calculated exactly.*

## 6 Numerical Evaluations

This section presents our simulation setup and numerical results. We use IEEE 9-bus, 14-bus, 30-bus, 39-bus, 57-bus, and 118-bus power systems in our evaluations [1]. All the experiments are implemented using MATLAB R2020a on a workstation with Intel(R) Xeon(R) W-2145 CPU with 3.70GHz processor and 128GB memory. Simulation codes can be found at [27].

IEEE Dataset	Reference Generators	Coherent Generator Groups (Using Bus Indices)	Maximum # of Iterations	Maximum Run time
9-Bus	1; 3	{1,2}, {3}	1	0.07 s
14-Bus	1; 6	{1:3}, {6,8}	7	0.31 s
30-Bus	1; 13; 22	{1,2}, {13}, {22,23,27}	6	0.43 s
39-Bus	30; 31; 37	{30}, {31:36}, {37:39}	9	1.11 s
57-Bus	1; 6; 9	{1:3}, {6,8}, {9,10}	21	5.23 s
118-Bus	10; 46; 49; 87	{10,12,25,26,31}, {46}, {49,54,59,61,65,66,69,80}, {87,89,100,103,111}	16	62.70 s

Table 1: First three columns show IEEE power system case study data used in the numerical evaluations. Last two columns present the maximum number of iterations and maximum run time that Algorithm 1 takes to converge. The maximum values in the last two columns are found across a set of experiments where the adversary budget ( $C$ ) is increased from  $C = 1$  until adversarial actions cause the power system to fail.

## 6.1 Simulation Setup

We extract the topology of power system, transmission line susceptance, load demands, generator capacities, transmission line capacities, and voltage angle bounds from the IEEE case study datasets [1]. The power flow in the system at the initial operating point is computed using Matpower [36]. The reference generators and generator coherent groups are chosen as in Table 1.

We initialize Algorithm 1 using adversary *pure* strategies that result in the four largest DC power flow disruptions to the system, a grid operator *pure* islanding strategy in the absence of any adversary (solution to optimization problem in Eqn. (19) with  $z_{i,j}^a = 1$  for all  $(i, j) \in \mathcal{L}$ ) and corresponding grid operator and adversary best responses, respectively.

We evaluate the performance of our approach by comparing the DC power flow disruption resulting from Algorithm 1 and a baseline case. A grid operator in the baseline case computes an islanding strategy without considering the presence of an adversary, i.e., the grid operator solves Eqn. (19) with  $z_{i,j}^a = 1$  for all  $(i, j) \in \mathcal{L}$ . The adversary in baseline observes the islanding strategy computed by the grid operator, and computes its best response by solving Eqn. (4). Let  $\mathcal{S}_{\mathbf{S}}$  and  $\mathcal{S}_{\mathbf{B}}$  denote the transmission lines tripped in the proposed model (Algorithm 1) and baseline case, respectively. We denote the DC power flow disruption corresponding to Algorithm 1 and baseline as  $R(\mathcal{S}_{\mathbf{S}})$  and  $R(\mathcal{S}_{\mathbf{B}})$ , respectively.

## 6.2 Case Study Results

Figure 1 illustrates the grid operator islanding strategy and adversary strategy obtained using Algorithm 1 on IEEE 39-bus data for adversary budget,  $C = 8$ . The grid operator performs islanding strategy 1 with probability (w.p.) 0.27 and islanding strategy 2 w.p. 0.73. In this case study we obtain  $R(\mathcal{S}_{\mathbf{B}}) \approx 10.04$  GW



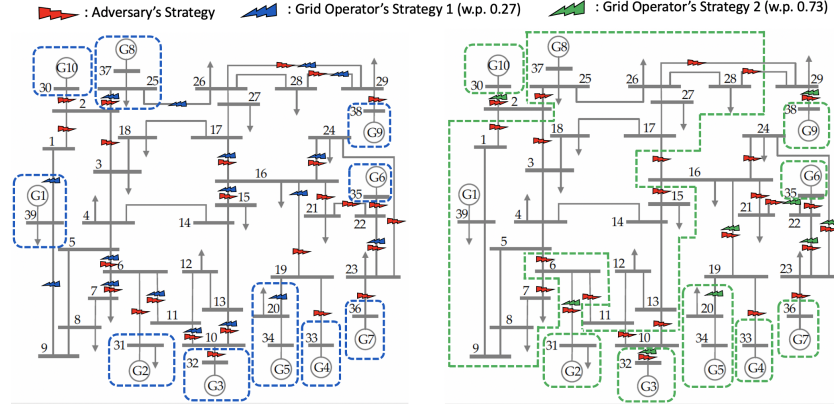


Fig. 1: Grid operator and adversary strategies obtained using Algorithm 1 on IEEE 39-bus data for adversary budget,  $C = 8$ . The islands induced by the grid operator and adversary strategies are marked by the dotted lines.

and  $R(\mathcal{S}_S) \approx 10.22$  GW. Hence, by committing to the islanding strategy given by Algorithm 1, the grid operator incurs  $\sim 180$  MW less DC power flow disruption.

Figure 2-(a) plots the reduced DC power flow disruption achieved by the grid operator via committing to an islanding strategy given by Algorithm 1 (i.e.,  $R(\mathcal{S}_B) - R(\mathcal{S}_S)$ ) for different values of adversary budget,  $C$ , under each test case given in Table 1. We construct a set of *attack scenarios* by increasing the values of  $C$  from  $C = 1$  until the value of  $C$  breaks down the grid (i.e., all the generators are isolated into individual islands). The results show that the grid operator achieves a better performance by committing to a strategy of Algorithm 1 under some attack scenarios and in other scenarios the grid operator achieves the same performance as committing to a baseline strategy. Algorithm 1 and the baseline achieve same performance when equilibrium strategies of the adversary and the grid operator do not contain any common set of transmission lines.

Figure 2-(b) shows the percentage of attack scenarios where the grid operator is able to achieve lower DC power flow disruption by committing to a strategy given by Algorithm 1. We only consider the attack scenarios that does not break down the grid when computing the related percentage values. The results suggest that on average grid operator is able to perform better in 44% of the attack scenarios and save 12.27 MW when committing to a strategy of Algorithm 1.

Last two columns of Table 1 present the maximum number of iterations and maximum run time of Algorithm 1 to converge across the attack scenarios considered under each case study. The results show that Algorithm 1 takes less than 21 iterations to converge for the cases analyzed. Also, run time to converge is less than 63 seconds in Algorithm 1 for the largest dataset (IEEE-118 bus) analyzed. For other cases, Algorithm 1 finds the optimal strategies in less than 5.23 seconds. Note that the worst-case number of iterations Algorithm 1 can take to converge is  $(2^L - 1)$ , (i.e., worst-case computation time is exponential in

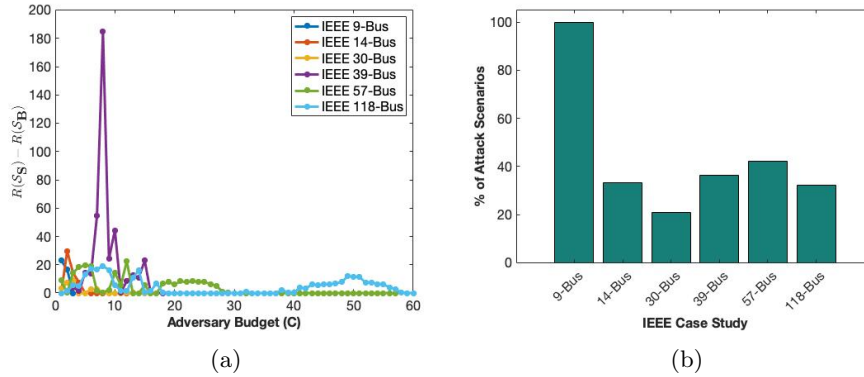


Fig. 2: Figure 2-(a) shows the reduced DC power flow disruption achieved by the grid operator via committing to an islanding strategy given by Algorithm 1 ( $R(\mathcal{S}_B)$ ) compared to a baseline case ( $R(\mathcal{S}_S)$ ). Figure 2-(b) shows the percentage of attack scenarios where the grid operator performs better by committing to a strategy given by Algorithm 1.

$L$ ). Therefore, the results suggest that Algorithm 1 converges with substantially less number of iterations compared with the worst-case bound.

## 7 Conclusion

In this paper, we studied the problem of controlled islanding of a power system in the presence of a malicious adversary. We formulated the interaction between the grid operator and adversary as a Stackelberg game. The grid operator first synthesizes a mixed strategy for controlled islanding, as well as the power generation for the post-islanding system. The adversary observes the islanding strategy of the grid operator. The adversary then compromises a subset of substations in the power system and trips the transmission lines that are connected with the compromised substations. We formulated an MINLP to compute the Stackelberg equilibrium of the game. To mitigate the computational challenge incurred by solving MINLP, we proposed a double oracle algorithm based approach to solve for the equilibrium strategies. The proposed approach solved a sequence of MILPs that model the best responses for both players. Additionally, we proved that the adversary's best response can be formulated as a submodular maximization problem under a cardinality constraint. We compared the proposed approach with a baseline, where the grid operator computes an islanding strategy by minimizing the power flow disruption without taking into account the adversary's response, using IEEE 9-bus, 14-bus, 30-bus, 39-bus, 57-bus, and 118-bus systems. The proposed approach outperformed the baseline in about 44% of test cases and saved about 12.27 MW power flow disruption on average.

## References

1. Illinois center for a smarter electric grid (icseg), <https://icseg.iti.illinois.edu/power-cases/>
2. Abur, A., Exposito, A.G.: Power system state estimation: theory and implementation. CRC press (2004)
3. Bosansky, B., Kiekintveld, C., Lisy, V., Pechoucek, M.: An exact double-oracle algorithm for zero-sum extensive-form games with imperfect information. *Journal of Artificial Intelligence Research* **51**, 829–866 (2014)
4. Chow, J.H.: Time-scale modeling of dynamic networks with applications to power systems, vol. 46. Springer (1982)
5. Conitzer, V.: On stackelberg mixed strategies. *Synthese* **193**(3), 689–703 (2016)
6. Delgadillo, A., Arroyo, J.M., Alguacil, N.: Analysis of electric grid interdiction with line switching. *IEEE Transactions on Power Systems* **25**(2), 633–641 (2009)
7. Demetriou, P., Asprou, M., Kyriakides, E.: A real-time controlled islanding and restoration scheme based on estimated states. *IEEE Transactions on Power Systems* **34**(1), 606–615 (2018)
8. Deng, R., Xiao, G., Lu, R.: Defending against false data injection attacks on power system state estimation. *IEEE Transactions on Industrial Informatics* **13**(1), 198–207 (2015)
9. Ding, L., Gonzalez-Longatt, F.M., Wall, P., Terzija, V.: Two-step spectral clustering controlled islanding algorithm. *IEEE Transactions on Power Systems* **28**(1), 75–84 (2012)
10. Haque, M., Rahim, A.: Identification of coherent generators using energy function. In: *IEE Proceedings C (Generation, Transmission and Distribution)*. vol. 137, pp. 255–260. IET (1990)
11. Hasan, S., Ghafouri, A., Dubey, A., Karsai, G., Koutsoukos, X.: Vulnerability analysis of power systems based on cyber-attack and defense models. In: *2018 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference*. pp. 1–5. IEEE (2018)
12. Henner, V.: A network separation scheme for emergency control. *International Journal of Electrical Power & Energy Systems* **2**(2), 109–114 (1980)
13. Jain, M., Korzhyk, D., Vaněk, O., Conitzer, V., Pěchouček, M., Tambe, M.: A double oracle algorithm for zero-sum security games on graphs. In: *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*. pp. 327–334 (2011)
14. Karwowski, J., Mańdziuk, J.: Double-oracle sampling method for Stackelberg equilibrium approximation in general-sum extensive-form games. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 34, pp. 2054–2061 (2020)
15. Krause, A., Golovin, D.: Submodular function maximization. *Tractability* **3**, 71–104 (2014)
16. Lee, M.R., Assante, J.M., Conway, T.: Analysis of the cyber attack on the Ukrainian power grid. [https://www.eisac.com/cartella/Asset/00006542/TLP\\_WHITE\\_E-ISAC\\_SANS\\_Ukraine\\_DUC\\_6\\_Modular\\_ICS\\_Malware%20Final.pdf?parent=64412](https://www.eisac.com/cartella/Asset/00006542/TLP_WHITE_E-ISAC_SANS_Ukraine_DUC_6_Modular_ICS_Malware%20Final.pdf?parent=64412)
17. Liu, J., Xiao, Y., Li, S., Liang, W., Chen, C.P.: Cyber security and privacy issues in smart grids. *IEEE Communications Surveys & Tutorials* **14**(4), 981–997 (2012)
18. Liu, Z., Clark, A., Bushnell, L., Kirschen, D.S., Poovendran, R.: Controlled islanding via weak submodularity. *IEEE Transactions on Power Systems* **34**(3), 1858–1868 (2018)

19. McMahan, H.B., Gordon, G.J., Blum, A.: Planning in the presence of cost functions controlled by an adversary. In: *Proceedings of the 20th International Conference on Machine Learning*. pp. 536–543 (2003)
20. Mo, Y., Kim, T.H.J., Brancik, K., Dickinson, D., Lee, H., Perrig, A., Sinopoli, B.: Cyber-physical security of a smart grid infrastructure. *Proceedings of the IEEE* **100**(1), 195–209 (2011)
21. Muir, A., Lopatto, J.: Final report on the August 14, 2003 blackout in the United States and Canada: causes and recommendations (2004)
22. Nedic, D.P., Dobson, I., Kirschen, D.S., Carreras, B.A., Lynch, V.E.: Criticality in a cascading failure blackout model. *International Journal of Electrical Power & Energy Systems* **28**(9), 627–633 (2006)
23. Paruchuri, P., Kraus, S., Pearce, J.P., Marecki, J., Tambe, M., Ordonez, F.: Playing games for security: An efficient exact algorithm for solving Bayesian Stackelberg games. In: *International Foundation for Autonomous Agents and Multiagent Systems* (2008)
24. Patsakis, G., Rajan, D., Aravena, I., Oren, S.: Strong mixed-integer formulations for power system islanding and restoration. *IEEE Transactions on Power Systems* **34**(6), 4880–4888 (2019)
25. Peiravi, A., Ildarabadi, R.: A fast algorithm for intentional islanding of power systems using the multilevel kernel k-means approach. *Journal of Applied Sciences* **9**(12), 2247–2255 (2009)
26. Pita, J., Jain, M., Ordóñez, F., Portway, C., Tambe, M., Western, C., Paruchuri, P., Kraus, S.: Using game theory for Los Angeles airport security. *AI Magazine* **30**(1), 43–43 (2009)
27. Sahabandu, D.: Controlled islanding code, <https://github.com/sdinuka/Controlled-Islanding-Code>
28. Sun, K., Zheng, D.Z., Lu, Q.: Splitting strategies for islanding operation of large-scale power systems using OBDD-based methods. *IEEE Transactions on Power Systems* **18**(2), 912–923 (2003)
29. Wang, W., Lu, Z.: Cyber security in the smart grid: Survey and challenges. *Computer networks* **57**(5), 1344–1371 (2013)
30. Yang, B., Vittal, V., Heydt, G.T., Sen, A.: A novel slow coherency based graph theoretic islanding strategy. In: *2007 IEEE Power Engineering Society General Meeting*. pp. 1–7. IEEE (2007)
31. Yang, Q., Yang, J., Yu, W., An, D., Zhang, N., Zhao, W.: On false data-injection attacks against power system state estimation: Modeling and countermeasures. *IEEE Transactions on Parallel and Distributed Systems* **25**(3), 717–729 (2013)
32. Yardley, J., Harris, G.: 2nd day of power failures cripples wide swath of India. <https://www.nytimes.com/2012/08/01/world/asia/power-outages-hit-600-million-in-india.html>
33. You, H., Vittal, V., Wang, X.: Slow coherency-based islanding. *IEEE Transactions on power systems* **19**(1), 483–491 (2004)
34. Zhao, Q., Sun, K., Zheng, D.Z., Ma, J., Lu, Q.: A study of system splitting strategies for island operation of power system: A two-phase method based on OBDDs. *IEEE Transactions on Power Systems* **18**(4), 1556–1565 (2003)
35. Zhu, Y., Yan, J., Sun, Y., He, H.: Revealing cascading failure vulnerability in power grids using risk-graph. *IEEE Transactions on Parallel and Distributed Systems* **25**(12), 3274–3284 (2014)
36. Zimmerman, R.D., Murillo-Sánchez, C.E., Thomas, R.J.: Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on power systems* **26**(1), 12–19 (2010)